

Figure 1

The novel gene as identified through RACE analysis (894 bp)

GGGAGTGGAGTGAGGGGTAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAA
AGCTTGCCGAACATAAGCAAGAATGTCTTGCTCGTGGTTTGGAGACCAAGGGAATAAAG
CAAGATCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAAT
GAAGAAGATGTACTGGGAGATGAAACAGAGGAAGAAGAAACAAAGCCCATTGAGCTCCC
TGTCAAAGAGGAAGAACCCCTGAAAAAACTGTTGATGTGGCAGCAGAGAAGAAAGTGG
TGAAAATTACATCTGAAATACCACAGACTGAGAGAATGCAGAAGAGGGCTGAACGATTCA
ATGTACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCT
TCAGTTCCAACAAAAGGTCTGTCTATCTGATAACAAACCTATGGTTAACTTGGATAAGCTG
AAGGAAAGAGCTCAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGAT
GATGAGAACTGAAAAAGAGGAAGGAGCGATTTGGGATTGTCACAAGTTCAGCTGGAAC
TGGAACACAGAGGATACAGAGGCAAAGAAGAGGAAAAGAGCAGAGCGCTTTGGGATT
GCCTGATGAAAAGTTCCTGATACTTTCTGTTCTCCAGTGTTTTCCATTTCTCTCCTTCTTC
TTGGTCACATATATGCCTAAATGCACAGTCATGTGCCTACGTCCTGCCTCGCAATGAGG
GAGCATGTACCCCAGGTACATCCATGAACTGCGGCAGCAGTTTGACTTATTGCTGTTTCA
GCTTTAAGGTTGTTGTGTTTTTGTGTTTTGATTATGTTGCTTGTTAATAAAAAAAAAATAGAAA

A

097846 0940
T0720 92488260

Figure 2

Amino acid sequence as translated from the novel gene (210 amino acids)

MATETVELHKLKLAELKQECLARGLETGKIQDLIHLQAYLEEHAEEDVVGDETEEE
ETKPIELPVKEEEPPEKTVDAAEKKVVKITSEIPQTERMQKRAERFNVPSLESKKAARAAR
FGISSVPTKGLSSDNKPMVNLDKLERARFGLNVSSISRKSEDDKLLKKRKERFGIVTSSAG
TGTTEDTEAKKRKRAERFGIA

Underlined sequences are amino acid sequences obtained by MS/MS analysis.

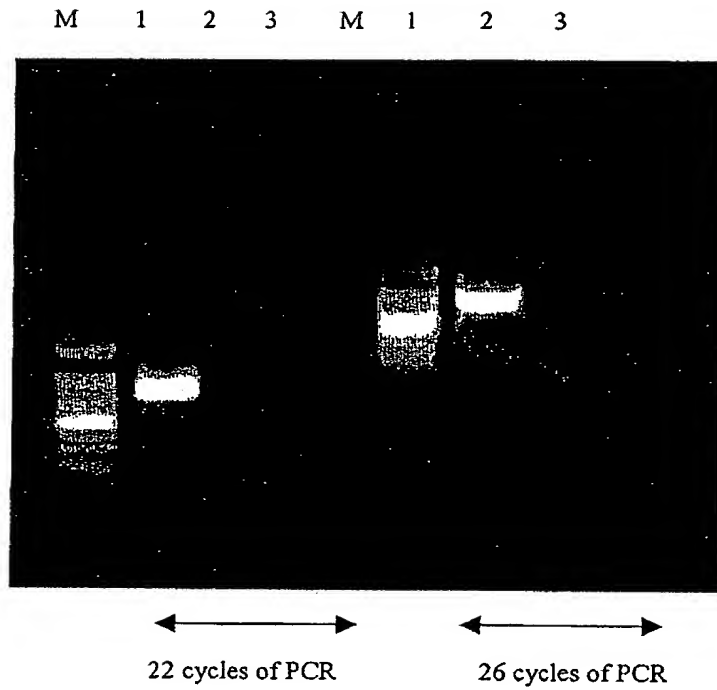
Figure 3

The sequence of the novel gene amplified through long distant PCR and used to construct the expression vector (873 bp).

TGGAGTGAGGGGTAACAAGATGGCGACCGAGACGGTGGAGCTCCATAAGCTAAAGCTT
GCCGA ACTAAAGCAAGAATGTCTTGCTCGTGGTTTGGAGACCAAGGGAATAAAGCAAGA
TCTTATCCACAGACTCCAGGCATATCTTGAAGAACATGCTGAAGAGGAGGCAAATGAAG
AAGATGTACTGGGAGATGAAACAGAGGAAGAAGAAACAAAGCCCATTGAGCTCCCTGTC
AAAGAGGAAGAACCCCTGAAAAAAGTGTGATGTGGCAGCAGAGAAGAAAGTGGTGAA
AATTACATCTGAAATACCACAGACTGAGAGAATGCAGAAGAGGGCTGAACGATTCAATGT
ACCTGTGAGCTTGGAGAGTAAGAAAGCTGCTCGGGCAGCTAGGTTTGGGATTTCTTCAG
TTCCAACAAAAGGTCTGTCTATCTGATAACAAACCTATGGTTAACTTGGATAAGCTGAAGG
AAAGAGCTCAAAGATTTGGTTTGAATGTCTCTTCAATCTCCAGAAAGTCTGAAGATGATG
AGAAACTGAAAAAGAGGAAGGAGCGATTTGGGATTGTCACAAGTTCAGCTGGAAGTGA
ACCACAGAGGATACAGAGGCAAAGAAGAGGAAAAGAGCAGAGCGCTTTGGGATTGCCT
GATGAAAAGTTCCTGATACTTTCTGTTCTCCAGTGTTTTCCATTTCTCTCCTTCTTCTTGG
TCACATATATGCCTAAATGCACAGTCATGTGCCTACGTCCTGCCTCGCAATGAGGGAGC
ATGTACCCAGGTACATCCATGAACTGCGGCAGCAGTTTGACTTATTGCTGTTTCAGCTT
TAAGGTTGTTGTGTTTTTGTGTTTGTATTATGTTGCTTGTTAAT

bioRxiv preprint doi: <https://doi.org/10.1101/2018.04.11.244444>; this version posted April 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4



09788475-022104
T07220-9488260

Figure 5

P-151 5'-Untranslated Region

1 75
CAGGGGCAGCAGTGATTATCTGAACTCGGATCTTTAAAATTGTGGTAGCTCTAAAGCTGATGATGTCTGGTTAGG

76 150
AAGTGGCTCTTGCCCGCCCCAGCCCCACCGCCAGTTCCTTAAGCCCGCCCCATGCCCTCCCAGCTTCCTCCTCA

151 225
TGTTTCATCGGTTTTTTCAGGGCTCCCTTCAACGCTCCCCTCTCAGTATTTAGGTCACCACTCCCTCGGCGCCCCCT

226 300
TTCGCCTCCCACCATTTTTCTCAGCAACCCTTACAGTCTTTGCAGCTCCTACCTGCCAGCTCAGATCCCCGTCC

301 375
GGCTATGGGGCGGGCGCCGGCTACCACACCTGAAGTCTCCAGGAAGTAACGCCTCTCCTTCTGCCCCCTTTCCTGT
376 450
TGGAGGAACAGAATCAGCGCTGCCACCACCCATTGGTTGGTGGTCTGTAATGCAGAAGCACAGTTGGTTGCCATT
451 525
TCTGTCTGTTTCGCAAGATACAGTGCCCGCCCCCTCTCCAGTTCCACCTTTTGAAGAGGTGGGGCAAGCTGCCTAG
526 600
AGAAGTGAGAGCGACGTCAGCTATTGACCAATGGGAAGAGCTGATGGTATGGCGTGGGAGCAAGAGTGACAACGA
601 675
TTGGTCAGCCTTGCATCTCTACGCCTAAGGCGGGAACCTCCTGGAGGCGGAGGCCGCGGGTGGGGGGAGTGGAGTG
676
AGGGGTAAACAAGATG.....P151 coding region.....

(Total length: 690 bp)

Sequence with asterisk: the 274 bp fragment

Underlined sequences are the minicistrones or uORFs before the start of the P151 coding region with the start and stop codons in bold.

P07220:347334..60